SEQUENCE LISTING

(1) GENERAL INFORMATION:

   (i) APPLICANT: Choi et. al.

   (ii) TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

   (iii) NUMBER OF SEQUENCES: 4

   (iv) CORRESPONDENCE ADDRESS:
      (A) ADDRESSEE: Human Genome Sciences, Inc.
      (B) STREET: 9410 Key West Avenue
      (C) CITY: Rockville
      (D) STATE: Maryland
      (E) COUNTRY: USA
      (F) ZIP: 20850

   (v) COMPUTER READABLE FORM:
      (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
      (B) COMPUTER: HP Vectra 486/33
      (C) OPERATING SYSTEM: MSDOS version 6.2
      (D) SOFTWARE: ASCII Text

   (vi) CURRENT APPLICATION DATA:
      (A) APPLICATION NUMBER: 08/961,083
      (B) FILING DATE:OCT-30-1997
      (C) CLASSIFICATION:

   (vii) PRIOR APPLICATION DATA:
      (A) APPLICATION NUMBER:60/029,960
      (B) FILING DATE:OCT-31-1996

   (viii) ATTORNEY/AGENT INFORMATION:
      (A) NAME: Michelle S. Marks
      (B) REGISTRATION NUMBER: 41,971
      (C) REFERENCE/DOCKET NUMBER: PB340P2
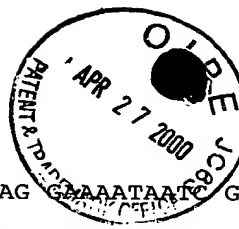
   (vi) TELECOMMUNICATION INFORMATION:
      (A) TELEPHONE: (301) 309-8504
      (B) TELEFAX: (301) 309-8512

(2) INFORMATION FOR SEQ ID NO: 1:

   (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 2389 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: double
      (D) TOPOLOGY: linear

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
TTCTTACGAG TTGGGACTGT ATCAAGCTAG AACGGTTAAG GAAAATAATC GTGTTTCCTA        60

TATAGATGGA AAACAAGCGA CGCAAAAAAC GGAGAATTTG ACTCCTGATG AGGTTAGCAA       120

GCGTGAAGGA ATCAATGCTG AGCAAATCGT CATCAAGATA ACAGACCAAG GCTATGTCAC       180

TTCACATGGC GACCACTATC ATTATTACAA TGGTAAGGTT CCTTATGACG CTATCATCAG       240

TGAAGAATTA CTCATGAAAG ATCCAAACTA TAAGCTAAAA GATGAGGATA TTGTTAATGA       300

GGTCAAGGGT GGATATGTTA TCAAGGTAGA TGGAAAATAC TATGTTTACC TTAAGGATGC       360

TGCCCACGCG GATAACGTCC GTACAAAAGA GGAAATCAAT CGACAAAAAC AAGAGCATAG       420

TCAACATCGT GAAGGTGGAA CTCCAAGAAA CGATGGTGCT GTTGCCTTGG CACGTTCGCA       480

AGGACGCTAT ACTACAGATG ATGGTTATAT CTTTAATGCT TCTGATATCA TAGAGGATAC       540

TGGTGATGCT TATATCGTTC CTCATGGAGA TCATTACCAT TACATTCCTA AGAATGAGTT       600

ATCAGCTAGC GAGTTGGCTG CTGCAGAAGC CTTCCTATCT GGTCGAGGAA ATCTGTCAAA       660

TTCAAGAACC TATCGCCGAC AAAATAGCGA TAACACTTCA AGAACAAACT GGGTACCTTC       720

TGTAAGCAAT CCAGGAACTA CAAATACTAA CACAAGCAAC AACAGCAACA CTAACAGTCA       780

AGCAAGTCAA AGTAATGACA TTGATAGTCT CTTGAAACAG CTCTACAAAC TGCCTTTGAG       840

TCAACGACAT GTAGAATCTG ATGGCCTTGT CTTTGATCCA GCACAAATCA CAAGTCGAAC       900

AGCTAGAGGT GTTGCAGTGC CACACGGAGA TCATTACCAC TTCATCCCTT ACTCTCAAAT       960

GTCTGAATTG GAAGAACGAA TCGCTCGTAT TATTCCCCTT CGTTATCGTT CAAACCATTG      1020

GGTACCAGAT TCAAGGCCAG AACAACCAAG TCCACAACCG ACTCCGGAAC CTAGTCCAGG      1080

CCCGCAACCT GCACCAAATC TTAAAATAGA CTCAAATTCT TCTTTGGTTA GTCAGCTGGT      1140

ACGAAAGTT GGGGAAGGAT ATGTATTCGA AGAAAAGGGC ATCTCTCGTT ATGTCTTTGC      1200

GAAAGATTTA CCATCTGAAA CTGTTAAAAA TCTTGAAAGC AAGTTATCAA AACAAGAGAG      1260

TGTTTCACAC ACTTTAACTG CTAAAAAAGA AAATGTTGCT CCTCGTGACC AAGAATTTTA      1320

TGATAAAGCA TATAATCTGT TAACTGAGGC TCATAAAGCC TTGTTTGNAA ATAAGGGTCG      1380

TAATTCTGAT TTCCAAGCCT TAGACAAATT ATTAGAACGC TTGAATGATG AATCGACTAA      1440

TAAAGAAAAA TTGGTAGATG ATTTATTGGC ATTCCTAGCA CCAATTACCC ATCCAGAGCG      1500

ACTTGGCAAA CCAAATTCTC AAATTGAGTA TACTGAAGAC GAAGTTCGTA TTGCTCAATT      1560

AGCTGATAAG TATACAACGT CAGATGGTTA CATTTTTGAT GAACATGATA TAATCAGTGA      1620

TGAAGGAGAT GCATATGTAA CGCCTCATAT GGGCCATAGT CACTGGATTG GAAAAGATAG      1680

CCTTTCTGAT AAGGAAAAAG TTGCAGCTCA AGCCTATACT AAAGAAAAAG GTATCCTACC      1740

TCCATCTCCA GACGCAGATG TTAAAGCAAA TCCAACTGGA GATAGTGCAG CAGCTATTTA      1800

CAATCGTGTG AAAGGGGAAA AACGAATTCC ACTCGTTCGA CTTCCATATA TGGTTGAGCA      1860

TACAGTTGAG GTTAAAAACG GTAATTTGAT TATTCCTCAT AAGGATCATT ACCATAATAT      1920
```

```
TAAATTTGCT TGGTTTGATG ATCACACATA CAAAGCTCCA AATGGCTATA CCTTGGAAGA      1980

TTTGTTTGCG ACGATTAAGT ACTACGTAGA ACACCCTGAC GAACGTCCAC ATTCTAATGA      2040

TGGATGGGGC AATGCCAGTG AGCATGTGTT AGGCAAGAAA GACCACAGTG AAGATCCAAA      2100

TAAGAACTTC AAAGCGGATG AAGAGCCAGT AGAGGAAACA CCTGCTGAGC CAGAAGTCCC      2160

TCAAGTAGAG ACTGAAAAAG TAGAAGCCCA ACTCAAAGAA GCAGAAGTTT TGCTTGCGAA      2220

AGTAACGGAT TCTAGTCTGA AAGCCAATGC AACAGAAACT CTAGCTGGTT TACGAAATAA      2280

TTTGACTCTT CAAATTATGG ATAACAATAG TATCATGGCA GAAGCAGAAA AATTACTTGC      2340

GTTGTTAAAA GGAAGTAATC CTTCATCTGT AAGTAAGGAA AAAATAAAC                  2389
```

(2) INFORMATION FOR SEQ ID NO: 2:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 796 amino acids
        (B) TYPE: amino acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: protein

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Ser Tyr Glu Leu Gly Leu Tyr Gln Ala Arg Thr Val Lys Glu Asn Asn
1               5                   10                  15

Arg Val Ser Tyr Ile Asp Gly Lys Gln Ala Thr Gln Lys Thr Glu Asn
            20                  25                  30

Leu Thr Pro Asp Glu Val Ser Lys Arg Glu Gly Ile Asn Ala Glu Gln
            35                  40                  45

Ile Val Ile Lys Ile Thr Asp Gln Gly Tyr Val Thr Ser His Gly Asp
        50                  55                  60

His Tyr His Tyr Tyr Asn Gly Lys Val Pro Tyr Asp Ala Ile Ile Ser
65                  70                  75                  80

Glu Glu Leu Leu Met Lys Asp Pro Asn Tyr Lys Leu Lys Asp Glu Asp
                85                  90                  95

Ile Val Asn Glu Val Lys Gly Gly Tyr Val Ile Lys Val Asp Gly Lys
            100                 105                 110

Tyr Tyr Val Tyr Leu Lys Asp Ala Ala His Ala Asp Asn Val Arg Thr
            115                 120                 125

Lys Glu Glu Ile Asn Arg Gln Lys Gln Glu His Ser Gln His Arg Glu
    130                 135                 140

Gly Gly Thr Pro Arg Asn Asp Gly Ala Val Ala Leu Ala Arg Ser Gln
145                 150                 155                 160

Gly Arg Tyr Thr Thr Asp Asp Gly Tyr Ile Phe Asn Ala Ser Asp Ile
            165                 170                 175
```

```
Ile Glu Asp Thr Gly Asp Ala Tyr Ile Val Pro His Gly Asp His Tyr
            180                 185                 190

His Tyr Ile Pro Lys Asn Glu Leu Ser Ala Ser Glu Leu Ala Ala Ala
        195                 200                 205

Glu Ala Phe Leu Ser Gly Arg Gly Asn Leu Ser Asn Ser Arg Thr Tyr
    210                 215                 220

Arg Arg Gln Asn Ser Asp Asn Thr Ser Arg Thr Asn Trp Val Pro Ser
225                 230                 235                 240

Val Ser Asn Pro Gly Thr Thr Asn Thr Asn Thr Ser Asn Asn Ser Asn
                245                 250                 255

Thr Asn Ser Gln Ala Ser Gln Ser Asn Asp Ile Asp Ser Leu Leu Lys
            260                 265                 270

Gln Leu Tyr Lys Leu Pro Leu Ser Gln Arg His Val Glu Ser Asp Gly
        275                 280                 285

Leu Val Phe Asp Pro Ala Gln Ile Thr Ser Arg Thr Ala Arg Gly Val
    290                 295                 300

Ala Val Pro His Gly Asp His Tyr His Phe Ile Pro Tyr Ser Gln Met
305                 310                 315                 320

Ser Glu Leu Glu Glu Arg Ile Ala Arg Ile Ile Pro Leu Arg Tyr Arg
                325                 330                 335

Ser Asn His Trp Val Pro Asp Ser Arg Pro Glu Gln Pro Ser Pro Gln
            340                 345                 350

Pro Thr Pro Glu Pro Ser Pro Gly Pro Gln Pro Ala Pro Asn Leu Lys
        355                 360                 365

Ile Asp Ser Asn Ser Ser Leu Val Ser Gln Leu Val Arg Lys Val Gly
    370                 375                 380

Glu Gly Tyr Val Phe Glu Glu Lys Gly Ile Ser Arg Tyr Val Phe Ala
385                 390                 395                 400

Lys Asp Leu Pro Ser Glu Thr Val Lys Asn Leu Glu Ser Lys Leu Ser
                405                 410                 415

Lys Gln Glu Ser Val Ser His Thr Leu Thr Ala Lys Lys Glu Asn Val
            420                 425                 430

Ala Pro Arg Asp Gln Glu Phe Tyr Asp Lys Ala Tyr Asn Leu Leu Thr
        435                 440                 445

Glu Ala His Lys Ala Leu Phe Xaa Asn Lys Gly Arg Asn Ser Asp Phe
    450                 455                 460

Gln Ala Leu Asp Lys Leu Leu Glu Arg Leu Asn Asp Glu Ser Thr Asn
465                 470                 475                 480

Lys Glu Lys Leu Val Asp Asp Leu Leu Ala Phe Leu Ala Pro Ile Thr
                485                 490                 495

His Pro Glu Arg Leu Gly Lys Pro Asn Ser Gln Ile Glu Tyr Thr Glu
            500                 505                 510
```

```
Asp Glu Val Arg Ile Ala Gln Leu Ala Asp Lys Tyr Thr Thr Ser Asp
        515                 520             525

Gly Tyr Ile Phe Asp Glu His Asp Ile Ile Ser Asp Glu Gly Asp Ala
    530                 535             540

Tyr Val Thr Pro His Met Gly His Ser His Trp Ile Gly Lys Asp Ser
545                 550             555                 560

Leu Ser Asp Lys Glu Lys Val Ala Ala Gln Ala Tyr Thr Lys Glu Lys
            565             570             575

Gly Ile Leu Pro Pro Ser Pro Asp Ala Asp Val Lys Ala Asn Pro Thr
        580             585             590

Gly Asp Ser Ala Ala Ala Ile Tyr Asn Arg Val Lys Gly Glu Lys Arg
        595             600             605

Ile Pro Leu Val Arg Leu Pro Tyr Met Val Glu His Thr Val Glu Val
    610             615             620

Lys Asn Gly Asn Leu Ile Ile Pro His Lys Asp His Tyr His Asn Ile
625             630             635                 640

Lys Phe Ala Trp Phe Asp Asp His Thr Tyr Lys Ala Pro Asn Gly Tyr
            645             650             655

Thr Leu Glu Asp Leu Phe Ala Thr Ile Lys Tyr Tyr Val Glu His Pro
            660             665             670

Asp Glu Arg Pro His Ser Asn Asp Gly Trp Gly Asn Ala Ser Glu His
        675             680             685

Val Leu Gly Lys Lys Asp His Ser Glu Asp Pro Asn Lys Asn Phe Lys
    690             695             700

Ala Asp Glu Glu Pro Val Glu Glu Thr Pro Ala Glu Pro Glu Val Pro
705             710             715                 720

Gln Val Glu Thr Glu Lys Val Glu Ala Gln Leu Lys Glu Ala Glu Val
            725             730             735

Leu Leu Ala Lys Val Thr Asp Ser Ser Leu Lys Ala Asn Ala Thr Glu
            740             745             750

Thr Leu Ala Gly Leu Arg Asn Asn Leu Thr Leu Gln Ile Met Asp Asn
        755             760             765

Asn Ser Ile Met Ala Glu Ala Glu Lys Leu Leu Ala Leu Leu Lys Gly
        770             775             780

Ser Asn Pro Ser Ser Val Ser Lys Glu Lys Ile Asn
785                 790             795
```

(2) INFORMATION FOR SEQ ID NO: 3:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 37 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AGTCGGATCC TTCTTACGAG TTGGGACTGT ATCAAGC   37

(2) INFORMATION FOR SEQ ID NO: 4:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 40 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AGTCAAGCTT GTTTATTTTT TCCTTACTTA CAGATGAAGG   40